

AGNPS WATERSHED MODELING WITH GIS DATABASES

**Michael P. Finn, Computer Programmer/Analyst; E. Lynn Usery, Research Geographer;
Douglas J. Scheidt, Computer Scientist Trainee; Thomas Beard, Geographer; Sheila Ruhl,
Cartographic Technician; Morgan Bearden, Cartographer**

U.S. Geological Survey
Mid-Continent Mapping Center
1400 Independence Road
Rolla, MO 65401
Phone: 573-308-3931
Fax: 573-308-3652
mfinn@usgs.gov

Abstract: Geospatial databases built with geographic information systems served as primary sources of input data to the Agricultural NonPoint Source pollution model of watershed hydrology. Elevation, land cover, and soil data for four watersheds are the base from which we extracted the 22 input parameters required by the Agricultural NonPoint Source model. The study demonstrates the utility of employing geospatial databases as sources in watershed modeling to investigate the improved accuracy of the water-quality model results, and it shows examples of the automatic extraction of model input parameters from these databases. The study examines the implications of the results for modelers using this model in four watersheds: two in Georgia, and one each in Indiana and Washington. To demonstrate the development and practical utility of the user-friendly interface, a tool was created for generating input parameters, executing the pollution model, and analyzing model output. This tool uses object-oriented programming and macro languages to manipulate the raster databases. This investigation demonstrates new methods of automatically extracting the requisite input parameters for the Agricultural NonPoint Source pollution model from geographic information systems databases.

INTRODUCTION

Designers of watershed models strive to provide decisionmakers with information (usually water quantity and water quality, particularly physical, biological, and chemical components of quality) about a watershed and the watershed's response to environmental factors. The U.S. Department of Agriculture (USDA), as the lead agency, developed the Agricultural NonPoint Source (AGNPS) pollution model of watershed hydrology in response to the complex problem of managing nonpoint sources of pollution. AGNPS simulates the behavior of runoff, sediment, and nutrient transport from watersheds that have agriculture as their prime use. The model operates on a cell basis and is a distributed parameter, event-based (one storm event) model. The model requires 22 input parameters (covering hydrologic, soils, drainage, agricultural management, and other information). Output parameters are grouped primarily by hydrology, sediment, and chemical output (Young *et*

Disclaimer: Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U. S. Government.

al., 1995).

Geospatial databases built with geographic information systems (GIS) served as primary sources of input data to AGNPS. Elevation, land cover, and soil data for four watersheds are the base from which we extracted the 22 input parameters required by the AGNPS. Our effort provides an example of automatic extraction of the AGNPS input parameters from high-resolution GIS databases to investigate the improved accuracy of the water-quality model results. To demonstrate this objective of automatic parameter extraction, we followed the general process of parameter extraction from the geospatial data through a computer program (the AGNPS Data Generator) specifically created to generate the pollution model required input. Generating the 22 input parameters required varying degrees of computational complexity that fell into three simplified categories: complex development, straightforward development (using, primarily, ERDAS Imagine's Spatial Modeler), and simple development (consisting, primarily, of "hard coded" values based on the advice of local experts).

STUDY AREAS

Four Watersheds: The four watersheds used include Little River and Piscola Creek, Georgia, Sugar Creek, Indiana, and EL68D Wasteway, Washington. Little River, Sugar Creek, and EL68D Wasteway were selected because they are National Water-Quality Assessment (NAWQA) Program sites with periodic sampling by USGS personnel. Piscola Creek was added because of previous work there and data availability. In the case of the Little River watershed, in addition to being a NAWQA site with data availability, it was selected because there have been many years of continuous sampling there. Little River is located in Tift, Turner, and Worth Counties and covers approximately 33,242 hectares. Piscola Creek is located in Brooks and Thomas Counties and covers approximately 44,414 hectares (Figure 1). Sugar Creek is located in Henry, Hancock, and Madison Counties and covers approximately 23,976 hectares (Figure 2). Except for its southernmost tip, which is in Franklin County, all of the EL386D Wasteway is in Adams County, within and surrounding (to the north, south, and east) Othello, Washington. EL68D Wasteway is approximately 37,719 hectares (Figure 3).

Watershed Boundaries: We used two sets of watershed boundaries for this study. The first set was the NAWQA established watershed boundaries. We used U.S. Geological Survey (USGS) digital elevation models (DEM) to extract the second set. The USGS collects and assesses information on water chemistry, hydrology, land use, and stream habitat in more than 50 major rivers across the Nation as a part of the NAWQA Program. Part of the program is concerned with water quality and nonpoint sources in agricultural watersheds (USGS, 2001). The GIS Weasel, a USGS computer program, produced the second set of boundaries using the DEMs. The GIS Weasel interfaces GIS software with several water models in the Modular Modeling System (Leavesley *et al.*, 2002). Because the NAWQA boundary does not usually match the flow according to the DEMs, due to resolution and accuracy issues, we used two sets of boundaries. Because the DEM determines the resulting GIS Weasel boundary, this boundary is consistent with the slope and other data derived from the DEM.

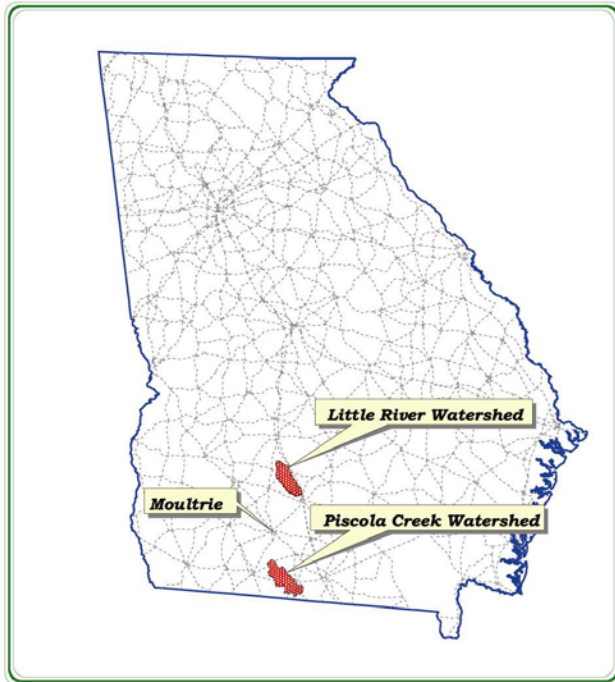


Figure 1. Little River and Piscola Creek Watersheds, Georgia

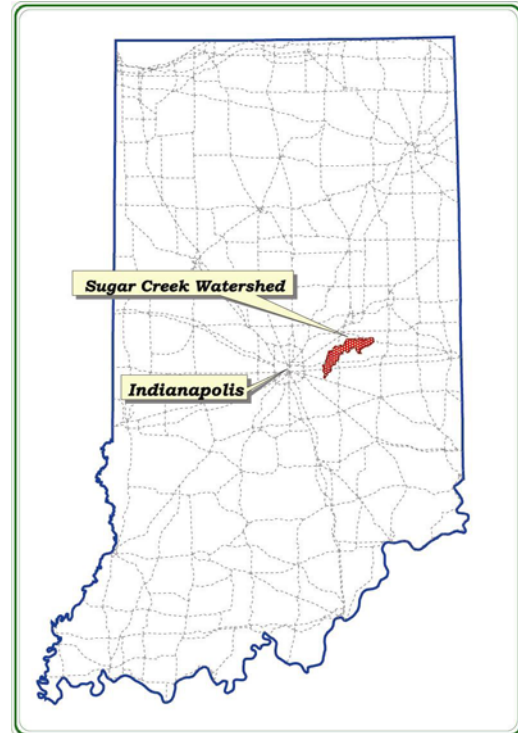


Figure 2. Sugar Creek Watershed, Indiana.



Figure 3. EL68D Wasteway Watershed, Washington.

GIS DATABASES FOR PARAMETER EXTRACTION

We used USGS 30-m DEMs and the 30-m National Land Characteristics Data land cover data as the base for the extraction processes. We augmented these land cover data with recent (1997 and 2001) Landsat thematic mapper data. In addition, we augmented these databases with a set of high-resolution (3-m) elevation and land cover data to help determine resolution effects. We created the soil databases from USDA soil surveys by scanning mylar separates of soil polygons, then rectifying, vectorizing, and tagging the resulting digital data. We resampled the soil data to the 30-m and 3-m base resolutions. To assess the effects of resolution on model results, we resampled the 30-m raster data to 60-, 120-, 210-, 240-, 480-, 960-, and 1,920-m resolution. The 210-m database roughly matches the 10-acre grid size commonly used by the USDA.

AGNPS PARAMETER GENERATION

AGNPS Data Generator: We created the AGNPS Data Generator computer program to provide a user-friendly interface between ERDAS Imagine and the AGNPS model (version 5.0). AGNPS Version 5.0 was written in the C programming language and was created from the source code from AGNPS version 4.03 (Young *et al.*, 1994; Witte *et al.*, 1995). It is important that users have an efficient method of dealing with the difficult process of generating the AGNPS requisite input parameters and a method for analyzing the data it produces. We designed and developed the entire interface for Imagine 8.4, running on WinNT/2000. We designed this graphical user interface (GUI) primarily to simplify the task of creating the input for AGNPS. Because of the lack of GIS-AGNPS interfaces available to users, we created this Data Generator program to fill that need, particularly with regard to personal computers running the Windows operating system. Figure 4 is a screen shot of the Data Generator that shows the concise interface for creating AGNPS input parameters, running AGNPS, and creating images of output for analysis.

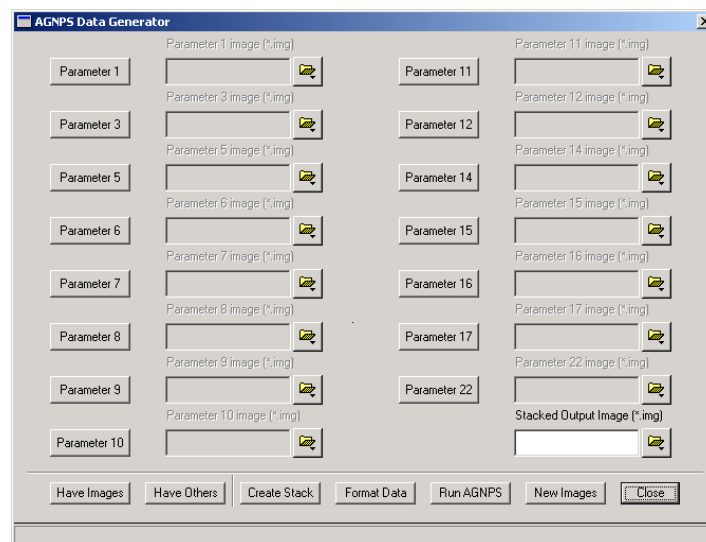


Figure 4. Screenshot of AGNPS Data Generator

We selected ERDAS Imagine to develop the Data Generator on its foundation because Imagine is designed to allow users to implement their own programs and because the AGNPS program processes the data on a grid, or raster, basis. The AGNPS program is a simple command line program using just a few arguments. The difficult part of executing AGNPS is creating and setting up all the data it requires. In addition, we designed the AGNPS Data Generator to manipulate AGNPS output data to aid in analysis by creating images that display the data visually. A user initiates the Data Generator from the Imagine menu bar. We designed the GUI to work at a screen resolution of 800 x 600 pixels or greater in a concise view. All buttons on the GUI display a window for creating that parameter. We wrote all the displays using the ERDAS Macro Language.

Input Parameter Generation: To extract parameters, we used the AGNPS Data Generator. In creating the 22 input parameters, we needed varying degrees of computational development. General categories of these parameters are complex development, straightforward development, and simple development. Table 1 summarizes the generation of the requisite input parameters for AGNPS; following the table is a discussion of those parameters.

Table 1. Parameter Generation

Number	Title	Information on Generation
1	Cell Number	Using a watershed cutout of the DEM to create the watershed cells with unique cell numbers from the DEM.
2	Cell Division	Set to zero. No cells were divided.
3	Receiving Cell Number	Calculated by using cell number and flow direction within Imagine Spatial Modeler.
4	Receiving Cell Division	Set to zero. No cells were divided.
5	Flow Direction	Created the TARDEM program, and then processing with Imagine Spatial Modeler to edit flow-direction values.
6	SCS Curve Number	Determined by Imagine's Spatial Modeler, using the soil information and land cover as cross-referencing lookup tables.
7	Average Land Slope	Calculated by using Imagine Spatial Modeler's PERCENT SLOPE function on the DEM.
8	Slope Shape Factor	Calculated by using cell number, flow direction, and land slope within Imagine Spatial Modeler.
9	Slope Length	Calculated by executing a model that uses land slope and a maximum slope length within Imagine Spatial Modeler.
10	Overland Manning's Coefficient	Created with Imagine Spatial Modeler by using land cover as a lookup table.
11	Soil Erodibility Factor	Created with Imagine Spatial Modeler by using soils as a lookup table.
12	Cropping Factor	Created with Imagine Spatial Modeler by using land cover as a lookup table.
13	Practice Factor	Set to one (1).
14	Surface Condition Constant	Created with Imagine Spatial Modeler by using land cover as a lookup table.
15	COD (Chemical Oxygen	Created with Imagine Spatial Modeler by using land cover as a

	Demand) Factor	lookup table.
16	Soil Type	Created with Imagine Spatial Modeler by using soils as a lookup table.
17	Fertilizer Level	Created with Imagine Spatial Modeler by using land cover as a lookup table.
18	Pesticide Type	Set to zero (0).
19	Number of Point Sources	Set to zero (0).
20	Additional Erosion Sources	Set to zero (0).
21	Number of Impoundments	Set to zero (0).
22	Type of Channel	Created by running DEM through stages of TARDEM program, and then through Imagine Spatial Modeler along with land cover.

Details on Generation of Parameters:

Cell Number: Parameter 1, Cell Number, uses a watershed cutout of the DEM at a specific raster cell size (i.e., 30-m grid cells) to assign a value to all of the cells in the watershed from 1 to n, where n is the last cell in the watershed. The numbering begins at the upper left cell, moves along a row until there are no more watershed cells, and then proceeds to the next row to continue the numbering. In this manner, each cell of the watershed is assigned a unique number that can be used to identify it.

Receiving Cell Number: Parameter 3, Receiving Cell Number, uses the cell number and flow direction to determine the cell into which the subject cell flows. For each cell, the model will use the direction of flow (using a unique 2^n notation) and look in the proper direction to determine the cell number of the one the cell flows into next (Figure 5). The algorithm accomplishes this by employing eight custom 3x3 matrices (representing the eight possible directions, each focused on a neighboring cell of the center cell). Once the direction of the flow is known for the center cell, the corresponding matrix, which matches that flow direction, will read the value (cell number) in that direction and store it in the output raster coverage.

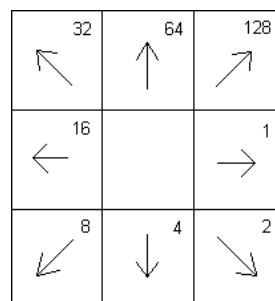


Figure 5. Flow Direction.

SCS Curve Number: Parameter 6, Soil Conservation Service (SCS) Curve Number, uses both the soil and the land cover images to resolve the correct curve number. A lookup is performed on the

soils attributes to find out the soil type (*i.e.*, A, B, C, or D). The algorithm then executes a pick function to determine the corresponding column in the land cover attributes table on the basis of the soil type.

Slope Shape Factor: Parameter 8, Slope Shape Factor, works essentially the same as parameter 3, except the algorithm uses slope values along with the cell number and flow direction. The relationship of the slope values at each point determines the slope shape factor (Figure 6). In addition to using flow direction and 3x3 matrices, the algorithm calculates Parameter 8 using the values of the center cell and both the cells in front of and behind it, this relationship being based on the direction of flow (*i.e.*, the cell flowing into the subject cell and the cell the subject cell flows into next). Once the algorithm determines the values for the slope of those three cells, it calculates the slope shape (Figure 7).










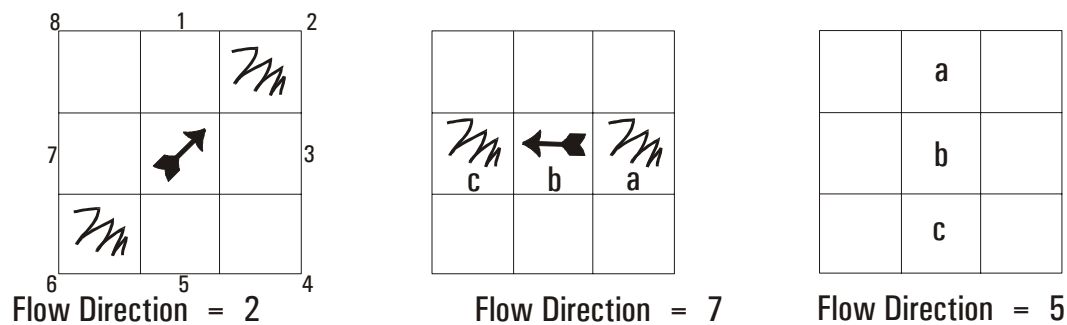
Slope Shape Factor:			a	b	c	<u>slope relation</u> (at each point)		
1 =		=				$a = b = c$		straight
2 =		=				$a < b < c$		convex
3 =		=				$a > b > c$		concave

Figure 6. Define Slope Shape Factor.

Determine Factor:

Given Flow Direction and Slope Percentage: Use flow direction of a cell "b" to compare the cell behind (slope "a") and the cell in front (having slope "c"). Compare all three slopes to form a relation 1, 2, or 3 for the slope shape factor.



if $a = b = c$, then Slope Shape Factor = 1
 if $a < b < c$, then Slope Shape Factor = 2
 if $a > b > c$, then Slope Shape Factor = 3

Figure 7. Determine Slope Shape Factor.

A straight slope would result when the slope percentages of three cells are equal (or nearly equal).

The AGNPS Data Generator assumes that anything within 0.1 percent is equal. A convex slope would occur when the tail slope is less than both the middle slope and the front slope, in addition to the middle slope being less than the front slope. A concave slope would occur in the opposite case, where the tail slope is greater than both slopes in front of it, and the middle slope is also greater than the front slope, as follows:

Where c is the front slope, b is the center cell slope, and a is the tail slope (Figure 7):

$a = b = c$ Straight (1)

$a < b < c$ Convex (2)

$a > b > c$ Concave (3)

Slope Length: Parameter 9, Slope Length, is currently a concern. It seems that there is no unambiguous, consensus method for determining slope length. Sample data that were provided with the AGNPS program contained the value of 100 ft for the slope length. Many discussions led to the belief that a maximum value should be 300 ft. Therefore, by taking the maximum, practical slope angle (45 degrees), one would be required to multiply the slope by roughly 6.6 in order to fit the AGNPS allowable range of 0-300. By taking the slope, fitting it in this range, and subtracting the new value from 300, we calculate the slope length. We have discussed various other approaches and algorithms but, on low-resolution data, the resulting slope length from these equations is too large to be meaningful.

Parameters 10, 11, 12, 14, 15, 16, and 17: The program, using Imagine's Spatial Modeler, creates these straightforward parameters (see Table 1 for parameter names) by simple lookups into the land cover or soil attributes.

Parameters 13, 18, 19, 20, and 21: The values for these simple parameters were hard coded on the basis of advice from experts (*e.g.* hydraulic, biological, or agricultural engineers) in the local area. (See Table 1 for parameter names). For example, Parameter 13, Practice Factor, is hard coded to 1. The factor is a ratio of soil loss to the corresponding loss with up-and-down-slope culture (Young *et al.*, 1994). For AGNPS input, this means that it is a worst-case situation.

Type of Channel: Parameter 22, Type of Channel, is created primarily by using a program suite called TARDEM, developed by David G. Tarboton (2000) of Utah State University. This collection of programs works with elevation data to create, among other things, a Strahler stream order. Other programs, such as Arc/Info and GIS Weasel, did not generate a high enough stream order to meet our requirements. The ASCII output from TARDEM served as input to a spatial model, which then reassigned the stream orders that were not large enough for concern. The stream order can then be imported to an image file and through the model be combined with land cover to make certain that all bodies of water are accounted for.

Extraction Methods: We extracted the requisite 22 AGNPS parameters from the three primary databases using object-oriented programming and macro languages embodied in our AGNPS Data Generator. The ERDAS Imagine software was the primary tool used for manipulating the raster GIS databases. The extracted parameters served as input to the event-based AGNPS for each of the watershed resolutions and for the two different watershed boundaries from NAWQA and Weasel.

CREATING AGNPS INPUT, OUTPUT, AND IMAGES

Input Data File Creation: After the input parameters are generated, the next step is to format these into a data file that AGNPS will accept and be able to read. For programming control and to meet our software design, we stacked Imagine created images in order on the basis of their parameter values into one image (“img” file). We processed this image using our AGNPS Data Generator program to extract the data and format the information into the proper order as an AGNPS data file. Using the Imagine Developer’s Toolkit, we extracted the information from the image, converted it from binary to decimal, and wrote it in the proper cell-by-cell orientation required by AGNPS. Using the data from the image, we added parameters to the input file within the AGNPS Data Generator, such as Soil Information, which is optional information that is required whenever no water cell is present. The values for additional parameters were hard coded into the program on the basis of value of the input.

Output Image Creation: We inspected the output in the AGNPS standard output tabular/ numerical form. In addition, we created graphical output for each model run by generating a series of multidimensional images from the numerical AGNPS output for each data resolution and model run. The program “agrun.exe” (as controlled by the AGNPS Data Generator) creates a nonpoint source (“nps”) file. This is simply a data file much like the input data for running AGNPS (Figure 8). Combining this information with the Parameter 1 (Cell Number) image allows the creation of new images so that the user can graphically display the output of AGNPS. The Data Generator uses the Parameter 1 file to get the correct geographic orientation of the output information for the watershed. In addition, the Data Generator uses Parameter 1 to gather statistics (so there is less need for user intervention) and to set the proper map model and projection information for the new output images.

The basic flow for creating new images is create an image with a specific number of layers, fill all of the layers with the data from the AGNPS output file, and set projection information and statistics for each image. The resulting images consist of multiple layers displaying the different runoff created by a single model event (Table 2). Table 2 lists the data (per band) for each image. The “xxx” at the beginning of the filenames in Table 2 represents the cell size. (For any resolution, the program creates images to display the AGNPS 5.0 output). For example, see Figures 9 and 10 for images created by the AGNPS Data Generator to aid users in visually analyzing model output. The user can choose to display these images in Imagine, where the values for multiple layers can be seen at one time for any particular cell of the watershed the user wishes to evaluate.

FEEDLOT

INITIAL

```

Watershed data
76055.88 10.89 7.30 160.00 7838 000 4.88 24618.76 2376.20
0.25 0.00 0.00 0.13 0.00 0.00 0.00 0.00
****
SEDIMENT
0.01 0.12 11 22 25.77 0.01 1082.05
0.01 0.03 13 9 9.91 0.01 416.20
0.10 0.03 3 1 8.36 0.00 350.95
0.13 0.08 3 1 9.63 0.01 404.43
0.38 0.02 0 0 2.92 0.00 122.56
0.64 0.23 4 1 56.58 0.03 2376.20
****
SOIL LOSS
1 000 21.78 2.32 6.23 112.34 0.00 0.00 0.0
0.00 0.61 0.03 0.00 100
0.00 0.38 0.03 0.00 100
0.02 1.99 0.22 0.00 100
0.03 0.46 0.28 0.00 100
0.08 0.14 0.83 0.00 100
0.13 3.57 1.39 0.00 100
2 000 98.01 5.54 3.20 208.72 0.00 0.00 0.0
0.00 1.16 0.00 0.00 100
0.00 0.42 0.00 0.00 100
0.00 0.34 0.00 0.00 100
0.00 0.18 0.00 0.00 100
0.00 0.05 0.00 0.00 100
0.00 2.15 0.00 0.00 100
3 000 10.89 4.75 0.00 0.00 0.00 0.00 0.0
0.11 0.00 1.17 0.00 100
0.11 0.00 1.17 0.00 100
0.86 0.00 9.40 0.00 100
1.08 0.00 11.75 0.00 100
3.24 0.00 35.24 0.00 100
5.39 0.00 58.74 0.00 100
.
.
.
NUTRIENT
1 000 21.78 0.52 0.00 0.59 0.00 0.00
0.26 0.00 0.03 0.00 0.00 34.11 0.00 0.00
2 000 98.01 0.00 0.00 1.42 0.00 0.00
0.00 0.00 0.06 0.00 0.00 0.00 0.00 0.00
3 000 10.89 10.35 0.00 3.23 0.00 0.00
5.18 0.00 0.60 0.00 0.00 182.83 0.00 0.00
4 000 10.89 1.09 0.00 1.05 0.00 0.00
0.54 0.00 0.07 0.00 0.00 51.12 0.00 0.00
5 000 10.89 0.54 0.00 0.59 0.00 0.00
.
.
.

```

Figure 8. AGNPS Data Output Example.

RESULTS

Collaboration continues between geographers, computer programmers, hydrologists, and hydraulic engineers to quantify the impact of geospatial resolution on model results. This study demonstrated the efficacy of using GIS databases as sources in watershed modeling, particularly with the AGNPS Pollution Model. We have demonstrated methods of automatically extracting the requisite input parameters for AGNPS from these databases. This study showed implications of the results for watershed modelers using the AGNPS model based on four study watersheds. Finally, we demonstrated the development and practical utility of an AGNPS-GIS Interface, the AGNPS Data Generator, as a tool for generating input, executing AGNPS, and analyzing model output.

Table 2.: AGNPS Data Generator Output Images (*.img) of AGNPS Version 5.00

Filename	Band	Definition	Units
xxxhydro	++		
	1	Drainage Area	acres
	2	Equivalent runoff for the cell (Overland Runoff)	inches
	3	Accumulated runoff volume into cell (Upstream Runoff)	inches
	4	Upstream Concentrated Flow (Peak Flow Upstream)	cfs
	5	Accumulated runoff volume out of cell (Downstream Runoff)	inches
	6	Downstream Concentrated Flow (Peak Flow Downstream)	cfs
	7	Runoff generated above cell	%
xxxclay	++		
	1	Eroded sediment (Cell Erosion)	tons/acre
	2	Upstream sediment yield	tons
	3	Sediment generated within cell	tons
	4	Sediment yield	tons
	5	Deposition in the cell	%
xxxsilt	++	Repeat for same variables as xxxclay	
xxxSAGG	++	Repeat for same variables as xxxclay	
xxxLAGG	++	Repeat for same variables as xxxclay	
xxxsand	++	Repeat for same variables as xxxclay	
xxxtotal	++	Repeat for same variables as xxxclay	
xxxnitro	++		
	1	Drainage area	acres
	2	Cell sediment nitrogen	lbs/acre
	3	Sediment attached nitrogen	lbs/acre
	4	Soluble nitrogen in cell runoff	lbs/acre
	5	Total soluble nitrogen	lbs/acre
	6	Soluble nitrogen concentration	ppm
xxxphospho	++		
	1	Cell sediment phosphorus	lbs/acre
	2	Sediment attached phosphorus	lbs/acre
	3	Soluble phosphorus in cell runoff	lbs/acre
	4	Total soluble phosphorus	lbs/acre
	5	Soluble phosphorus conc.	ppm
	6	Cell COD yield	lbs/acre
	7	Total soluble COD	lbs/acre
	8	Soluble COD concentration	ppm

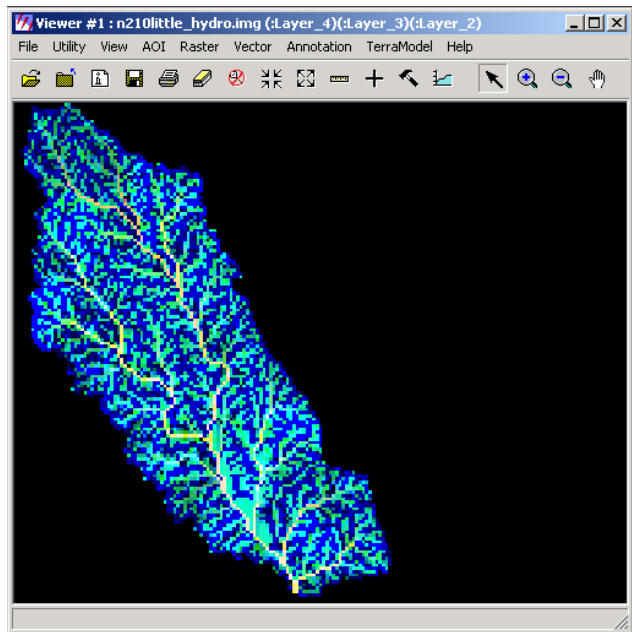


Figure 9. Data Generator Image of Hydrology Output. Band: Red, 4, Upstream concentrated flow; Green, 3, Accumulated runoff volume into cell; Blue, 2, Equivalent runoff by cell.

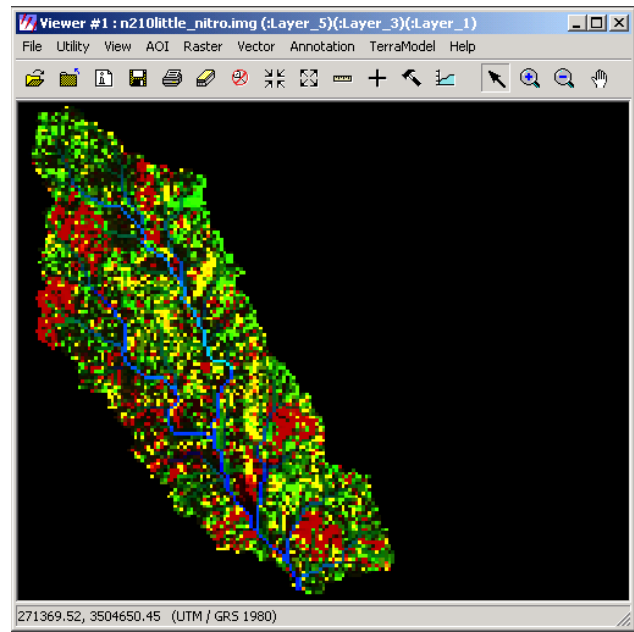


Figure 10. Data Generator Image of Nitrogen Output. Band: Red, 1, Drainage area; Green, 3, Sediment attached nitrogen; Blue, 5, Total soluble nitrogen.

REFERENCES

- Leavesley, G.H., Markstrom, S.L., Restrepo, P.J., Viger, R.J., in press, A modular approach to addressing model design, scale and parameter estimation issues in distributed hydrological modeling: Hydrological Processes.
- Tarboton, David G., 2000. TARDEM, A Suite of Programs for the Analysis of Digital Elevation Data. Internet at <http://www.engineering.usu.edu/dtarb/tardem.html>.
- U.S. Geological Survey, 2001. The National Water-Quality Assessment Program – Informing water-resource management and protection decisions. Internet at <http://water.usgs.gov/nawqa/docs/xrel/external.relevance.pdf>
- Witte, John, Theurer, Fred D., Baker, Kevin D., 1995. AGNPS Version 5.00 Verification: Software. AGNPS Web Site. Internet at <http://www.sedlab.olemiss.edu/AGNPS.html>
- Young, R.A., Onstad C.A., and Bosch, D.D., 1995. AGNPS: An Agricultural NonPoint Source Model. In Singh, Vijay P., Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado.
- Young, Robert A., Onstad, Charles A., Bosch, David D., Anderson Wayne P., 1994. AGricultural Non-Point Source Pollution Model, Version 4.03, AGNPS User's Guide. North Central Soil Conservation Research Laboratory, Morris, Minnesota.